Title: Automated Generation and Ensemble-Learned Matching of X-ray Absorption Spectra

C. Zheng[a], K. Mathew[b], C. Chen[a], Y. Chen[a], H. Tang[a], A. Dozier[c], J.J. Kas[d], F.D. Vila[d], J.J. Rehr[d], L.F.J. Piper[e,f], K. Persson[b], and S. P. Ong[a]
a) Dept. of NanoEngineering, Univ. of California San Diego, 9500 Gilman Dr, Mail Code 0448, La Jolla, CA 92093-0448, USA, b) Dept. of Materials Science, Univ. of California Berkeley, Berkeley, CA 94720, USA, c) Div. of Applied Research and Technology, National Institute for Occupational Safety and Health, Centers for Disease Control, Cincinnati, OH 45226, USA, d) Dept of Physics, Univ. of Washington, Seattle, WA 98195, USA, e) Dept. of Physics, Applied Physics and Astronomy, Binghamton Univ., Binghamton, NY 13902, USA, f) Materials Science & Engineering, Binghamton Univ., Binghamton, NY 13902, USA.
chz022@eng.ucsd.edu

X-ray absorption Spectroscopy (XAS) is a widely-used materials characterization technique. Despite its importance, only very sparse XAS reference resources are available to the community. Publicly available spectroscopic database such as EELS Database[1] contains 271 spectra, but only 21 of which are XAS spectra that covers 8 elements in the periodic table. In this talk, we will present the development of XAS database (XASdb), a large computed XAS database with more than 800,000 K-edge X-ray absorption near-edge spectra (XANES), and a novel Ensemble-Learned Spectra IdEntification (ELSIE) algorithm that leverages ensemble learning techniques to identify similar XANES spectra from our computed reference XASdb.[2] We will demonstrate the development of a high-throughput automation framework for X-ray absorption fine structure (XAFS) calculations which combines the power of the Python Materials Genomics (pymatgen) materials analysis library,[3] the FireWorks workflow management software[4] and FEFF[5]. With more than 800,000 K-edge XANES for over 40,000 unique materials, XASdb constitutes the largest existing collection of computed XAS spectra to date. The computed spectra in the XASdb have been made available in the Materials Project website (https://materialsproject.org/). A new web application – the XASApp – has been developed which allows any user to compare multiple X-ray absorption spectra and find matches within the XASdb. We believe the combination of the XASdb with these machine-learned spectra matching tools will be an invaluable resource to the materials research community by greatly enhancing the efficiency at which experimental XAS spectra can be analyzed.

1. Ewels P, Sikora T, Serin V, Ewels CP, Lajaunie L. *Microsc Microanal* 2017; : 717–724.
2. Zheng C, *et al. ArXiv e-prints* 2017.http://arxiv.org/abs/1711.02227.
3. Ong SP, *et al. Comput Mater Sci* 2013; **68**: 314–319.
4. Jain A, *et al. Concurr Comput Pract Exp* 2015; **27**: 5037–5059.
5. Rehr JJ, Kas JJ, Vila FD, Prange MP, Jorissen K. *Phys Chem Chem Phys* 2010; **12**: 5503.

This work is supported by the National Science Foundation's Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21) program under Award No. 1640899. The Materials Project, supported by the Department of Energy (DOE) Basic Energy Sciences (BES) program, under Grant No. EDCBEE is gratefully acknowledged for web dissemination and data infrastructure. The FEFF project is supported primarily by DOE BES Grant DE-FG02-97ER45623.